

U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE		ATTORNEY'S DOCKET NUMBER
FORM PTO-1390 (REV 5-93)		RUT 00-0010
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		U.S. APPLICATION NO. (If known, see 37 CFR 1.5) not yet assigned 09/762105
INTERNATIONAL APPLICATION NO. PCT/US99/17806	INTERNATIONAL FILING DATE 3 August 1999	PRIORITY DATE CLAIMED 3 August 1998 *
TITLE OF INVENTION TRANSLATION CONTROL ELEMENTS FOR HIGH-LEVEL PROTEIN EXPRESSION IN THE PLASTIDS OF HIGHER PLANTS AND METHODS OF USE THEREOF		
APPLICANT(S) FOR DO/EO/US MALIGA, Pal; KURODA, Hiroshi; KHAN, Muhammad		
Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information.		
<p>1. <input checked="" type="checkbox"/> This is a FIRST submission of items concerning a filing under 35 U.S.C. 371.</p> <p>2. <input type="checkbox"/> This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371.</p> <p>3. <input checked="" type="checkbox"/> This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).</p> <p>4. <input checked="" type="checkbox"/> A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.</p> <p>5. <input checked="" type="checkbox"/> A copy of the International Application as filed (35 U.S.C. 371(c)(2)) a. <input type="checkbox"/> is transmitted herewith (required only if not transmitted by the International Bureau). b. <input checked="" type="checkbox"/> has been transmitted by the International Bureau. c. <input type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US)</p> <p>6. <input type="checkbox"/> A translation of the International Application into English (35 U.S.C. 371(c)(2)).</p> <p>7. <input type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)) a. <input type="checkbox"/> are transmitted herewith (required only if not transmitted by the International Bureau). b. <input type="checkbox"/> have been transmitted by the International Bureau. c. <input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired. d. <input type="checkbox"/> have not been made and will not be made.</p> <p>8. <input type="checkbox"/> A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).</p> <p>9. <input type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).</p> <p>10. <input type="checkbox"/> A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).</p>		
Items 11. to 16. below concern other document(s) or information included:		
<p>11. <input type="checkbox"/> An Information Disclosure Statement under 37 CFR 1.97 and 1.98.</p> <p>12. <input type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.</p> <p>13. <input type="checkbox"/> A FIRST preliminary amendment. <input type="checkbox"/> A SECOND or SUBSEQUENT preliminary amendment.</p> <p>14. <input type="checkbox"/> A substitute specification.</p> <p>15. <input type="checkbox"/> A change of power of attorney and/or address letter.</p> <p>16. <input type="checkbox"/> Other items or information:</p>		
<p>Copy of Form PCT/IB/308 Copy of PCT General Power of Attorney Form</p> <p>*3 August 1998 15 December 1998 29 April 1999 11 June 1999</p> <p>Copy of Form PCT/IB/304</p>		

U.S. APPLICATION NO. 10/097762105 not yet assigned		INTERNATIONAL APPLICATION NO PCT/US99/17806	ATTORNEY'S DOCKET NUMBER RUT 00-0010																
<p>17. <input checked="" type="checkbox"/> The following fees are submitted:</p> <p>Basic National Fee (37 CFR 1.492(a)(1)-(5)): Search Report has been prepared by the EPO or JPO.....</p> <p>International preliminary examination fee paid to USPTO (37 CFR 1.482)</p> <p>No international preliminary examination fee paid to USPTO (37 CFR 1.482) but international search fee paid to USPTO (37 CFR 1.445(a)(2))..</p> <p>Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO.....</p> <p>International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(2)-(4).....</p>		<p>CALCULATIONS <small>PTO USE ONLY</small></p>																	
<p>ENTER APPROPRIATE BASIC FEE AMOUNT = \$ 690.00</p> <p>Surcharge of \$130.00 for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input checked="" type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(e)).</p>		<p>\$ 130.00</p>																	
<table border="1"> <thead> <tr> <th>Claims</th> <th>Number Filed</th> <th>Number Extra</th> <th>Rate</th> </tr> </thead> <tbody> <tr> <td>Total Claims</td> <td>28 -20 -</td> <td>8</td> <td>X 18 \$ 144.00</td> </tr> <tr> <td>Independent Claims</td> <td>5 -3 -</td> <td>2</td> <td>X 78 \$ 156.00</td> </tr> <tr> <td colspan="2">Multiple dependent claims(s) (if applicable)</td> <td>+</td> <td>\$</td> </tr> </tbody> </table>		Claims	Number Filed	Number Extra	Rate	Total Claims	28 -20 -	8	X 18 \$ 144.00	Independent Claims	5 -3 -	2	X 78 \$ 156.00	Multiple dependent claims(s) (if applicable)		+	\$	<p>TOTAL OF ABOVE CALCULATIONS = \$ 1,120.00</p>	
Claims	Number Filed	Number Extra	Rate																
Total Claims	28 -20 -	8	X 18 \$ 144.00																
Independent Claims	5 -3 -	2	X 78 \$ 156.00																
Multiple dependent claims(s) (if applicable)		+	\$																
<p>Reduction by 1/2 for filing by small entity, if applicable. Verified Small Entity statement must also be filed. (Note 37 CFR 1.9, 1.27, 1.28).*NO LONGER REQUIRED*</p>		<p>\$ 560.00</p>																	
<p>SUBTOTAL = \$ 560.00</p>																			
<p>Processing fee of \$130.00 for furnishing the English translation later the <input type="checkbox"/> 20 <input checked="" type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(f)).</p>		<p>+ \$ 0</p>																	
<p>TOTAL NATIONAL FEE = \$ 560.00</p>																			
<p>Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). \$40.00 per property + \$ 0</p>																			
<p>TOTAL FEES ENCLOSED = \$ 560.00</p>																			
		<p>Amount to be: refunded \$ charged \$</p>																	
<p>a. <input checked="" type="checkbox"/> A check in the amount of \$ 560.00 to cover the above fees is enclosed.</p> <p>b. <input type="checkbox"/> Please charge my Deposit Account No. _____ in the amount of \$ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.</p> <p>c. <input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. <u>04-1406</u>. A duplicate copy of this sheet is enclosed.</p>																			
<p>NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.</p>																			
<p>SEND ALL CORRESPONDENCE TO:</p> <p>Kathleen D. Rigaut, Ph.D., J.D. DANN, DORFMAN, HERRELL AND SKILLMAN 1601 Market Street, Suite 720 Philadelphia, Pennsylvania 19103 United States of America</p>																			
<p><i>Kathleen D. Rigaut</i> SIGNATURE Kathleen D. Rigaut, Ph.D., J.D. NAME 43,047 REGISTRATION NUMBER</p>																			

Translation Control Elements for High-level Protein
Expression in the Plastids of Higher Plants and Methods
of Use Thereof

5

10 This application claims priority from United States
Provisional Applications 60/095,163, filed August 3,
1998, 60/112,257, filed December 15, 1998, 60/095,167
filed August 3, 1998, 60/131,611, filed April 29, 1999
and 60/138,764, filed June 11, 1999 under 35 U.S.C.
15 §119(e). The entire disclosures of each of the
foregoing are incorporated by reference herein.

20 Pursuant to 35 U.S.C. §202(c) it is acknowledged
that the U.S. Government has certain rights in the
invention described herein, which was made in part with
funds from the National Science Foundation, Grant Number
MCB-96-30763.

FIELD OF THE INVENTION

25 This invention relates to the fields of transgenic
plants and molecular biology. More specifically, the
invention provides vectors targeting the plastid genome
which contain translation control elements facilitating
high levels of protein expression in the plastids of
30 higher plants. Both monocots and dicots are
successfully transformed with the DNA constructs
provided herein.

BACKGROUND OF THE INVENTION

35 Several publications are referenced in this
application in order to more fully describe the state of

the art to which this invention pertains. The disclosure of each of these publications is incorporated by reference herein.

The chloroplasts of higher plants accumulate individual components of the photosynthetic machinery as a relatively large fraction of total cellular protein. The best example is the enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco) involved in CO₂ fixation which can make up 65% of the total leaf protein (Ellis, R.J. 1979). Because of the potentially attainable high protein levels, there is significant interest in exploring chloroplasts as an alternative system for protein expression. To date, protein levels expressed from transgenes in chloroplasts are below the levels of highly-expressed chloroplast genes. Highest levels reported thus far in leaves are as follows: 1% of neomycin phosphotransferase (Carrer et al., 1993); 2.5% β -glucuronidase (Staub and Maliga, 1993) and 3-5% of *Bacillus thuringiensis* (Bt) crystal toxins (McBride et al., 1995). An alternative system, based on a nuclear-encoded, plastid-targeted T7 RNA polymerase may offer higher levels of protein expression (McBride et al., 1994), although this yield may come at a price.

In bacteria, the rate limiting step of protein synthesis is usually the initiation of translation, involving the binding of the initiator tRNA (formyl-methionyl-tRNA_f) and mRNA to the 70S ribosome, recognition of the initiator codon, and the precise phasing of the reading frame of the mRNA. Translation initiation depends on three initiation factors (IF1, IF2, IF3) and requires GTP. The 30S subunit is guided to the initiation codon by RNA-RNA base pairing between the 3' of the 16S rRNA and the mRNA ribosome binding site,

or Shine-Dalgarno (SD) sequence, located about 10 nucleotides upstream of the translation initiation codon (Voorma, 1996). RNA-RNA interaction between the "downstream box" (DB), a 15 nt sequence downstream of the AUG translational initiation codon and complementary sequences in the 16S rRNA 3' sequence or anti-downstream box (ADB; nucleotide positions 1469-1483) may also facilitate loading of the mRNA onto the 30S ribosome subunit (Sprengart et al., 1996). In addition, specific protein-RNA interactions may also facilitate translation initiation (Voorma, 1996).

Key components of the prokaryotic translation machinery have been identified in plastids, including homologues of the bacterial IF1, IF2 and IF3 initiation factors and an S1-like ribosomal protein (Stern et al., 1997). Most plastid mRNAs (92%) contain a ribosome binding site or SD sequence: GGAGG, or its truncated tri- or tetranucleotide variant. This sequence is similar to the bacterial SD consensus 5'-UAAGGAGGUGA-3' (Voorma, 1996). High level expression of foreign genes of interest in the plastids of higher plants is extremely desirable. The present invention provides novel genetic translational control elements for use in plastid transformation vectors. Incorporation of these elements into such vectors results in protein expression levels comparable to those observed for highly expressed chloroplast genes in both monocots and dicots.

SUMMARY OF THE INVENTION

5' genetic regulatory regions contain promoters with distinct DNA sequence information which facilitates recognition by the RNA polymerase and translational control elements which facilitate translation. Both of

these components act together to drive gene expression.

In accordance with the present invention, chimeric 5' regulatory regions have been constructed which incorporate translation control elements. Incorporation of these chimeric 5' regulatory regions into plastid transforming vectors followed by transformation of target plant cells gives rise to dramatically enhanced levels of protein expression. These chimeric 5' regulatory regions may be used to advantage to express foreign genes of interest in a wide range of plant tissues. It is an object of the present invention to provide DNA constructs and methods for stably transforming plastids of multicellular plants containing such promoters.

In one embodiment of the invention recombinant DNA constructs for expressing at least one heterologous protein in the plastids of higher plants are provided. The constructs comprise a 5' regulatory region which includes a promoter element, a leader sequence and a downstream box element operably linked to a coding region of said at least one heterologous protein. The chimeric regulatory region acts to enhance translational efficiency of an mRNA molecule encoded by said DNA construct. Vectors comprising the DNA constructs are also contemplated in the present invention. Exemplary DNA constructs of the invention include the following chimeric regulatory regions: PrnnLatpB+DBwt, PrrnLatpB-DB, PrrnLatpB+DBm, PrrnLclpP+DBwt, PrrnclpP-DB, PrrnLrbcl+DBwt, PrrnLrbcl-DB, PrrnLrbcl+DBm, PrrnLpsbB+DBwt, PrrnLpsbB-DB, PrrnLpsbA+DBwt, PrrnLpsbA-DB, PrrnLpsbA-DB(+GC), PrrnLT7g10+DB/Ec, PrrnLT7g10+DB/pt, and PrrnLT7g10-DB. Downstream box sequences preferred for use in the constructs of the

invention have the following sequences:

5' TCCAGTCACTAGCCCTGCCTCGGCA'3 and 5' CCCAGTCATGAATCACA
AAGTGGTAA'3.

The 5' regulatory segments of the invention have been successfully employed to drive the expression of the bar gene from *S. hydroscopicus* in the plastids of higher plants. Synthetic bar genes have also been generated and expressed using the DNA constructs of the present invention. These constructs have been engineered to maximize transgene containment in plastids by incorporating rare codons into the coding region that are not preferred for protein translation in microorganisms and fungi.

In yet another embodiment of the invention, at least one fusion protein is produced utilizing the DNA constructs of the invention. An exemplary fusion protein has a first and second coding region operably linked to the 5' regulatory regions described herein such that production of said fusion protein is regulated by said 5' regulatory region. In one embodiment the first coding region encodes a selectable marker gene and the second coding region encodes a fluorescent molecule to facilitate visualization of transformed plant cells. Vectors comprising a DNA construct encoding such a fusion protein are also within the scope of the present invention. An exemplary fusion protein consists an aadA coding region operably linked to a green fluorescent protein coding region. These moieties may be linked by peptide linkers such as ELVEGKLELVEGLKVA and ELAVEGKLEVA.

Plasmids for transforming the plastids of higher plants, are also included in the present invention. Exemplary plasmids are selected from the group

consisting of pHK30 (B), pHK31 (B), pHK60, pHK32 (B), pHK33 (B), pHK34 (A), pHK35 (A), pHK64 (A), pHK36 (A), pHK37 (A), pHK38 (A), pHK39 (A), pHK40 (A), pHK41 (A), pHK42 (A), pHK43 (A), pMSK56, pMSK57, pMSK48, pMSK49, 5 pMSK35, pMSK53 and pMSK54.

Transgenic plants, both monocots and dicots harboring the plasmids set forth above are also contemplated to be within the scope of the invention.

In yet another embodiment of the invention, methods 10 are provided for producing transplastomic monocots. One method comprises a) obtaining embryogenic cells; b) exposing said cells to a heterologous DNA molecule under conditions whereby said DNA enters the plastids of said cells, said heterologous DNA molecule encoding at 15 least one exogenous protein, said at least one exogenous protein encoding a selectable marker; c) applying a selection agent to said cells to facilitate sorting of untransformed plastids from transformed plastids, said cells containing transformed plastids surviving and 20 dividing in the presence of said selection agent; d) transferring said surviving cells to selective media to promote plant regeneration and shoot growth; and e) rooting said shoots, thereby producing transplastomic monocot plants. The heterologous DNA molecule may be 25 introduced into the plant cell via a process selected from the group consisting of biolistic bombardment, Agrobacterium-mediated transformation, microinjection and electroporation. In one embodiment of the above described method, protoplasts are obtained from the 30 embryogenic cells and the heterologous DNA molecule is delivered to said protoplasts by exposure to polyethylene glycol. Suitable selection agents for the practice of the methods of the invention are

streptomycin, and paromomycin. Monocot plants which may be transformed using the methods of the invention include but are not limited to maize, millet, sorghum, sugar cane, rice, wheat, barley, oat, rye, and turf grass.

5 In a preferred embodiment a method for producing transplastomic rice plants is provided. This method entails the following steps: a) obtaining embryogenic calli; b) inducing proliferation of calli on modified
10 CIM medium; c) obtaining embryogenic cell suspensions of said proliferating calli in liquid AA medium; d) bombarding said embryogenic cells with microprojectiles coated with plasmid DNA; e) tranferring said bombarded cells to selective liquid
15 AA medium; f) transferring said cells surviving in AA medium to selective RRM regeneration medium for a time period sufficient for green shoots to appear; and g) rooting said shoots in a selective MS salt medium.

20 Plasmids suitable for transforming rice as set forth above include pMSK35 and pMSK53, pMSK54 and pMSK49. Transplastomic rice plants so produced are also contemplated to be within the scope of the invention.

25 In yet a final embodiment of the invention methods for containing transgenes in transformed plants are provided. An emplary method includes the following steps: a) determining the codon usage in said plant to be transformed and in microbes found in association with said plant; and b) genetically engineering said transgene sequence via the introduction of rare
30 microbial codons to abrogate expression of said transgene in said plant associated microbe. In an exemplary embodiment of the method described immediately above the transgene is a bar gene and said rare codons

are arginine encoding codons selected from the group consisting of AGA and AGG, and transgene is not expressed in *E.coli*.

5 The following definitions will facilitate the understanding of the subject matter of the present invention:

10 **Heteroplasmic**: refers to the presence of a mixed population of different plastid genomes within a single plastid or in a population of plastids contained in plant cells or tissues.

15 **Homoplasmic**: refers to a pure population of plastid genomes, either within a plastid or within a population contained in plant cells and tissues.

20 Homoplasmic plastids, cells or tissues are genetically stable because they contain only one type of plastid genome. Hence, they remain homoplasmic even after the selection pressure has been removed, and selfed progeny are also homoplasmic. For purposes of the present invention, heteroplasmic populations of genomes that are functionally homoplasmic (i.e., contain only minor populations of wild-type DNA or transformed genomes with sequence variations) may be referred to herein as "functionally homoplasmic" or "substantially 25 homoplasmic." These types of cells or tissues can be readily purified to a homoplasmic state by continued selection.

Plastome: the genome of a plastid.

Transplastome: a transformed plastid genome.

30 **Transformation of plastids**: stable integration of transforming DNA into the plastid genome that is transmitted to the seed progeny of plants containing the transformed plastids.

5 Selectable marker gene: the term "selectable marker gene" refers to a gene that upon expression confers a selective advantage to the plastids and a phenotype by which successfully transformed plastids or cells or tissues carrying the transformed plastid can be identified.

10 Transforming DNA: refers to homologous DNA, or heterologous DNA flanked by homologous DNA, which when introduced into plastids becomes part of the plastid genome by homologous recombination.

15 Operably linked: refers to two different regions or two separate genes spliced together in a construct such that both regions will function to promote gene expression and/or protein translation.

20 The detailed description as follows provides examples of preferred methods for making and using the DNA constructs of the present invention and for practicing the methods of the invention. Any molecular cloning and recombinant DNA techniques not specifically described are carried out by standard methods, as generally set forth, for example in Sambrook et al., "DNA Cloning, A Laboratory Manual," Cold Spring Harbor Laboratory, 1989.

25 BRIEF DESCRIPTION OF THE DRAWINGS

30 Figure 1A. Plastid mRNAs and the small (16S) ribosomal RNA contain complementary sequences downstream of AUG implicating interactions between mRNA and 16S rRNA during translation initiation in plastids. Proposed model is based on data in *E. coli* (Sprengart et al., 1996); for sequence of 16S rRNA see ref. (Shinozaki et al., 1986b). SD, Shine-Dalgarno sequence; ASD, anti SD region; DB, downstream box; ADB, anti DB region. Watson-

Crick (line) and G-U (closed circle) pairing are marked.

Figure 1B. Sequence of the anti-downstream-box regions (ADB sequence underlined) of the 16S rRNA in plastids (pt; this application) and in *E. coli* (Ec; Sprengart et al., 1996). The *E. coli* ADB box contains sequences between nucleotides 1469-1483 of the 16S rRNA (Sprengart et al., 1996), corresponding to nucleotides 1416-1430 of the tobacco 16S rRNA (Dams et al., 1988; sequence between nucleotides 104173-104187 in Shinozaki et al., 1986).

Figure 2A. Base-pairing between plastid ADB and *atpB*, *clpP*, *rbcL*, *psbB* and *psbA* mRNAs (underlined). Multiple alternative DB-ADB interactions are shown.

Nucleotides changed to reduce or alter mRNA-rRNA interaction are in lower case. The number of potential nucleotide pairs formed with the 26 nt ADB region is in parenthesis. The number of pairing events affected by mutagenesis is in bold.

Figure 2B. Complementarity of Prrn T7 phage gene 10 leader derivatives with the *E. coli* and plastid ADB sequences. Nucleotides changed to reduce or alter mRNA-rRNA interaction are in lower case. The number of potential nucleotide pairs formed with the 26 nt ADB region is in parenthesis.

Figure 3A. DNA sequence of the chimeric Prrn plastid promoter fragments with *atpB* and *clpP* translation control regions. The plasmid name that is the source of the promoter fragment is given in parenthesis. The Prrn promoter sequence is underlined; nucleotide at which transcription initiates in tobacco plastids is marked with filled circle; translational

initiation codon (ATG) is in bold; SD is underlined with a wavy line; nucleotides of the 5' and 3' restriction sites and point mutations are in lower case.

5 **Figure 3B.** DNA sequence of the chimeric Prrn plastid promoter fragments with *rbcL* and *psbB* translation control regions. For details see description of Fig. 3A.

10 **Figure 3C.** DNA sequence of the chimeric Prrn plastid promoter fragments with *psbA* translation control regions. For details see description of Fig. 3A.

15 **Figure 3D.** DNA sequence of the chimeric Prrn plastid promoter fragments with the T7 phage gene 10 (PrrnLT7g10+DB/Ec) plastid (PrrnLT7g10+DB/pt) and synthetic DB (PrrnLT7g10-DB). For details see description of Fig. 3A.

20 **Figure 4A.** Plastid transformation vector pPRV111A with chimeric *neo* genes. Plasmid serial numbers, for example pHK34, designate pPRV111A plastid transformation vectors derivatives; adjacent plasmid numbers in parenthesis (e.g. pHK14) designate the source of the chimeric *neo* gene in pUC118 or pBSIIKS+ vectors. Arrows mark orientation of the selectable marker gene (*aadA*) and of the chimeric *neo* gene. Plastid targeting sequences are underlined in bold. Components of the chimeric *neo* genes are: Prrn, rRNA operon promoter fragment; L, leader sequence; DB, downstream box; *Nhe*I site which serves as a synthetic DB is marked by a heavy line; *neo*, neomycin phosphotransferase coding region; 30 Tr*bcl*, *rbcL* 3'-untranslated region. 16SrDNA, *trnV*, *rps12/7* are plastid genes (Shinozaki et al., 1986). The restriction sites marked for: EcoRI, SphI, StuI, SacI, *Nhe*I, *Nco*I, *Xba*I, *Hind*III, *Bam*HI and *Bgl*III. Restriction

sites in brackets were eliminated during construction. The *neo* translation initiation in plasmid pHK36 is included in *Nco*I site (not marked). The presence and relative order of *Nhe*I (**) and *Nco*I (*) restriction sites in the plasmid pPRV111A -DB derivatives (pHK35, pHK37, pHK40, pHK42, pHK43) are marked by asterisks. The promoter sequences are shown in Figures 3B, C and D.

Figure 4B. Plastid transformation vector pPRV111B with chimeric neo genes. See description of Fig. 4A. The promoter sequences are shown in Fig. 3A.

Figure 5. Construction of Prrn promoter-plastid leader fragments by overlap extension PCR.

15 **Figure 6.** Construction by the PCR of
PrrnLT7g10+DB/Ec promoter (SacI-NheI fragment) in
plasmid pHK18.

Figure 7. Construction by PCR of the PrrnLT7g10+DB/pt promoter (SacI-NheI fragment) in plasmid pHK19.

Figure 8. Restriction map of plasmids pHK2 and pHK3 with the P_{rrn(L)} rbcL(S)::neo::TrbcL gene. Restriction enzyme cleavage sites are marked for: BamHI, EcoRI, HindIII, NcoI, NheI, SacI, XbaI.

Figure 9. DNA sequence of the Prrn(L) rbcL(S)::neo::TrbcL gene in plasmid pHK3. Plasmid pHK2 carries an identical neo gene, except that there is an EcoRI site upstream of the SacI site.

Figure 10. NPTII accumulation in tobacco leaves

detected by protein gel blot analysis. Amount of total soluble leaf protein (μ g) loaded on SDS-PAGE gel is indicated above the lanes. Lanes are designated with plasmid used for plant transformation; μ g protein loaded per lane is given below. NPTII standard and Nt-pTNH32 extracts were run as positive controls; extracts from wild-type non-transformed plants (wt) were used as negative controls.

10 **Figure 11.** The levels of neo mRNA in the transplastomic leaves. The blots were probed for neo (top) and cytoplasmic 25S rRNA as loading control (bottom). Positions of the monocistronic neo mRNA in vector pPRV111A (Figure 4A), the monocistronic neo and dicistronic neo-aadA transcripts in vector pPRV111B (Figure 4B) and the monocistronic neo and dicistronic *rbcL*-neo transcripts in pTNH32 transformed plants (Carrer et al., 1993) are marked. Lanes are designated with the transgenic plant serial number. 4 μ g total cellular RNA was loaded per lane.

20 **Figure 12.** Fraction of a codon encoding a particular amino acid and triplet frequency per 1000 codons in the mutagenized *atpB* and *rbcL* DB region.

25 Altered nucleotides are in lower case.

30 **Figure 13A.** NPTII accumulation in tobacco roots detected by protein gel blot analysis. Lanes are designated with the plasmid used for plant transformation; μ g protein loaded per lane is given below. NPTII standard was run as positive control; extracts from wild-type non-transformed plants (wt) were used as negative controls.

5 **Figure 13B.** Steady-state levels of neo mRNA in tobacco roots. The neo probe detects a monocistronic mRNA in plants transformed with vector pPRV111A (Figure 4A), and a monocistronic neo and a dicistronic neo-aadA transcript in plants transformed with vector pPRV111B (Figure 4B). Lanes are designated with the transgenic plant serial number. 4 μ g total cellular RNA was loaded per lane.

10 **Figure 14.** Protein gel blot analysis to detect NPTII accumulation in tobacco seeds. Lanes are designated with plasmid used for plant transformation; μ g protein loaded per lane is given below. NPTII standard was run as positive control; extracts from 15 wild-type non-transformed plants (wt) were used as negative controls.

20 **Figure 15A.** Diagram showing integration of the chimeric neo and aadA genes into the plastid genome by 25 two homologous recombination events via the plastid targeting sequences (underlined). On top is shown a diagram of plasmids pHK30 and pHK32 are plastid transformation vector pPRV111B derivatives (Zoubenko et al., 1994). Horizontal arrows mark gene orientation. For 30 description of chimeric neo genes, see Figure 4B. 16SrDNA, trnV, rps12/7 are plastid genes (Shinozaki et al., 1986). The restriction sites marked for: EcoRI (E), SacI (S), NheI (N), XbaI (X), HindIII (H), BamHI (Ba) and BglII Restriction sites in brackets were eliminated during construction. In the middle the wild-type plastid DNA region (Wt-ptDNA) targeted for insertion is shown. Lines connecting plasmids and ptDNA mark sites of homologous recombination at the end of the vector

plastid-targeting regions. The transformed plastid genome segment (T-ptDNA) map is shown on the bottom.

Figure 15B. DNA gel blot analysis confirms of integration of the *neo* and *aadA* genes into the plastid genome. The blot on top was probed with the plastid targeting sequence (Probe 1 in Figure 15A). It lights up 4.2-kb and 1.4-kb fragments in transplastomic lines, and a 3.1-kb fragment in wild-type (see Figure 15A). Note that the 1.4-kb signal is weak in most clones. The blot on the bottom was probed for *neo* sequences, which are present only in the transplastomic lines.

Figure 16A. Diagram showing integration of the *bar* gene into the tobacco plastid genome. Map of the plastid targeting region in plasmid pJEK6 is shown on top. The targeted region of the wild-type plastid genome (wt-ptDNA) is shown in the middle. Integrated transgenes in the transplastome (T-ptDNA) are shown at the bottom. Map positions are shown for: the *bar* gene; *aadA*, the selectable spectinomycin resistance gene; *16SrDNA* and *rps12/7*, plastid genes (Shinozaki et al., 1986). Arrows indicate direction of transcription. Map position of the probe (2.5 kb) is marked by a heavy line; the wild-type (2.9-kb) and transgenic (3.3-kb, 1.9-kb) fragments generated by *Sma*I and *Bgl*II digestion are marked by thin lines.

Figure 16B. DNA gel blot confirms integration of *bar* into the tobacco plastid genome. Data are shown for transplastomic lines Nt-pJEK6-2A through E, Nt-pJEK6-5A through E and Nt-pJEK6-13A and B, and the wild-type parental line. *Sma*I-*Bgl*II digested total cellular DNA was probed with the 2.5-kb *Apa*I-*Bam*HI plastid targeting sequence marked with heavy line in Figure 16A.

Figure 17. PAT assay confirms bar expression in tobacco plastids. PAT activity was determined by conversion of PPT into acetyl-PPT using radiolabeled ¹⁴C-Acetyl-CoA. Data are shown for transplastomic lines Nt-pJEK6-2D, Nt-pJEK6-5A and Nt-pJEK6-13B, nuclear transformant Nt-pDM307-10 and wild-type (wt).
5

Figure 18A. Transplastomic tobacco plants are herbicide resistant. Wild-type and pJEK6-transformed 10 plants 13 days after Liberty spraying (5 ml, 2% solution).

Figure 18B. Maternal inheritance of PPT resistance in the seed progeny. Seeds from reciprocal crosses with Nt-pJEK6-5A plants germinated on 0, 10 and 50 mg/L PPT. 15 wt x pJEK6-5A, transplastomic used as pollen parent; pJEK6-5A x wt, transplastomic line female parent. Resistant seedlings are green on PPT medium, sensitive seedlings are bleached.

Figure 19. The engineered bacterial bar coding 20 region DNA sequence in plasmid pJEK3 and pJEK6 and encoded amino acid sequence. Nucleotides encoding the *rbcL* five N-terminal amino acids are in lower case. Nucleotides added at the 3' end during construction are also in lower case. NcoI, BglII and XbaI cloning sites 25 are marked.

Figure 20A. The synthetic bar gene DNA sequence and the encoded amino acid sequence. The arginines encoded by AGA/AGG codons are in bold. Original nucleotides are 30 in capital letters, altered bases are in lower case. Restriction sites used for cloning are marked.

Figure 20B. The synthetic s2-bar gene DNA sequence and the encoded amino acid sequence. The arginines

encoded by AGA/AGG codons are in bold. Original nucleotides are in capital letters, altered bases are in lower case. Restriction sites used for cloning are marked.

5

Figure 21. Synthetic and bacterial *bar* genes. The *bar* coding region is expressed in the Prrn/TrbCL cassettes. Note that the Prrn promoters differ with respect to the translational control region.

10

Figure 22A. PAT is expressed in *E. coli* from *bar*, but not from *s-bar* coding region. PAT activity was determined by conversion of PPT into acetyl-PPT using radiolabeled ^{14}C -Acetyl-CoA. Data are shown for *E. coli* transformed with plasmids pJEK6 and pKO12 carrying the *bar* gene, and pKO8, carrying *s-bar*.

15

Figure 22B. PAT assay confirms expression of *bar* and *s-bar* in tobacco plastids. PAT activity was determined by conversion of PPT into acetyl-PPT using radiolabeled ^{14}C -Acetyl-CoA. Data are shown for transplastomic lines Nt-pJEK6-13B and Nt-pKO3-24a, B carrying *bar* and *s-bar*, respectively.

20

25

Figure 23A. Plastid transformation vector with FLARE16-S as selectable marker targeting the plastid inverted repeat region. DNA and protein sequence at the *aadA-gfp* junction. Nucleotides derived from *aadA* and *gfp* are in capital, adapters sequences and the point mutation used to create the *Bst*XI restriction site (bold) are in lower case.

30

Figure 23B. Physical map of plastid transformation vector with FLARE16-S as selectable marker targeting the plastid inverted repeat region. Shown are: the promoter

(P) and 3'UTR (T) of the *aadA16pt-gfp* coding region and its component parts (*aadA* and *gfp* coding regions); *rrn16* and *rps12/7* plastid genes; restriction endonuclease sites HindIII (removed), SpeI, XbaI, NcoI, BstXI, NheI, 5 EcoRI. In plasmid pMSK56 *aadA16pt-gfp* is expressed from the P_{rrn}:LatpBDB promoter and encodes FLARE16-S1. In plasmid pMSK57 *aadA16pt-gfp* is expressed from the P_{rrn}:LrbcLDB promoter and encodes FLARE16-S2.

10 **Figure 24.** Localization of FLARE16-S to tobacco
plastids by laser scanning confocal microscopy in
heteroplastomic tissue. Images were processed to detect
FLARE16-S (green) and chlorophyll fluorescence (red) and
both in a merged view. Sections are shown from plants
15 expressing FLARE16-S1 (a,b) and FLARE16-S2 (3c-f). Note
wild-type and transformed plastids in leaves (3a,c,d),
chromoplasts of petals (3b), trichomes (3e) and non-
green root plastids (f). White arrows mark
transplastomic organelles. Bars represent 25 μ m.

20 **Figure 25.** Immunoblot analysis of FLARE16-S
accumulation in chloroplasts. The amount of loaded
protein (μ g) is indicated above the lanes.
Quantification of FLARE16-S1 (Nt-pMSK56 plants) and
25 FLARE16-S2 (Nt-pMSK57 plants) is based on comparison
with a purified GFP dilution series. Extract from a
wild-type plant (Nt) was used as negative control.

30 **Figure 26A.** Amplification of border fragments
confirms integration of FLARE-S genes into the plastid
genome. Maps of the plastid targeting regions of the
rice (pMSK49) and tobacco (pMSK57) vectors, the segment
of the rice and tobacco plastid genomes targeted by the

vectors (Os-wt and Nt-wt), and the same regions after integration of FLARE-S genes. The ends of plastid targeting regions are connected with cognate sequences in the wild-type plastid genome. Plastid genes 16SrDNA, 5 *trnV* and *rps12/7* are marked only in the wild-type plastid genomes. The position of PCR primers (O1-O6) and the PCR fragments generated by them are also shown.

Figure 26B. Amplification of border fragments confirms integration of FLARE-S genes into the plastid genome. Gels with PCR-amplified left and right border fragments, and with *aadA* fragment. Results are shown for rice (Os-pMSK49-1 and Os-pMSK49-2) and tobacco (Nt-pMSK57) transplastomic lines and wild-type (Os-wt) rice. The molecular weight markers is EcoRI- and HindIII-digested λ DNA.

Figure 27. Localization of FLARE11-S3 to rice chloroplasts in the Os-pMSK49-5 line by laser scanning confocal microscopy. Images were processed to detect FLARE11-S (green) and chlorophyll fluorescence (red) and both in a merged view. Arrows point to mixed populations of plastids in cells. Bar represents 25 μ m.

Figure 28. The sequence of FLARE16-S is shown.

Figure 29. The sequence of FLARE16-S1 is shown.

Figure 30. The sequence of FLARE16-S2 is shown.

Figure 31. The sequence of FLARE11-S is shown.

Figure 32. The sequence of FLARE11-S3 is shown.

Figures 33A and 33B. The sequence of pMSK35 is shown.

5 **Figures 34A and 34B.** The sequence of pMSK49 is shown.

Figure 35. A table describing the FLARE constructs of the invention.

10 **DETAILED DESCRIPTION OF THE INVENTION**

DNA cassettes for high level protein expression in plastids are provided herein. Higher plant plastid mRNAs contain sequences within 50 nt downstream of AUG that are complementary to the 16S rRNA 3-region. These complementary sequences are approximately at the same position as DB sequences in *E. coli* mRNAs. See Figures 1A and 2A. Interestingly, the tentative plastid DB sequence significantly deviates from the *E. coli* DB consensus, since the tobacco plastid and *E. coli* 16S rRNA sequence in the anti-downstream-box (ADB) region is significantly different (Figure 1B). The feasibility of improving protein expression by incorporating DB sequences in plastids was assessed by constructing a series of chimeric 5' regulatory regions consisting of the plastid rRNA operon σ^{70} -type promoter (Prrn-114; Svab and Maliga, 1993; Vera and Sugiura, 1995) and the leader sequence of plastid mRNAs with the native DB, mutagenized DB and synthetic DB sequences. The plastid mRNA leaders differ with respect to the presence and position of the SD sequence. Translation efficiency from the chimeric promoters was determined by expressing the bacterial neo gene in plastids. The neo (or kan) gene encodes neomycin phosphotransferase (NPTII) and

confers resistance to kanamycin in bacteria and plastids (Carrer et al., 1993). We have found that NPTII from the chimeric *neo* transcripts accumulates in the range of 0.2% to 23% of the total soluble leaf protein,
5 indicating the importance of translational control signals in the mRNA 5' region for high-level protein expression.

There is great interest in producing recombinant proteins in plants plastids which, thus far have been expressed from nuclear genes only (Arntzen, 1997; Conrad and Fiedler, 1998; Kusnadi et al., 1997). Protein levels produced from the PrrnLrbcl+DBwt and PrrnLT7g10 expression cassettes described here significantly exceed protein levels reported for nuclear genes. Accumulation of NPTII from nuclear genes is typically <<0.1% (Allen et al., 1996), the highest value being 0.4% of the total soluble protein (Houdt et al., 1997). We reported earlier accumulation of 1% NPTII from a plastid *neo* transgene (Carrer et al., 1993). Other examples for protein accumulation from plastid transgenes are 2.5% β -glucuronidase (GUS) (Staub and Maliga, 1993)) and 3-5% of the *Bacillus thuringiensis* (Bt) crystal toxins (McBride et al., 1995). As compared to this earlier report, we have achieved a significant increase in NPTII levels, up to 23% of total soluble protein.
10
15
20
25

FLARE-S, a protein obtained by fusing an antibiotic-inactivating enzyme with the *Aequorea victoria* green fluorescence protein accumulated to 8% and 18% of total soluble protein from the PrrnLatpB+DBwt and PrrnLrbcl+DBwt cassettes provided herein. See Example 8. High-level protein accumulation from the cassettes of the present invention can be clearly attributed to engineering the translational control
30

region (TCR) of the chimeric genes. These novel genetic elements may be used in different applications to drive expression of proteins with agronomic, industrial or pharmaceutical importance.

5 There is a strong demand for methods that control the flow of transgenes in field crops. Incorporation of the transgenes in the plastid genome rather than the nuclear genome results in natural transgene containment, since plastids are not transmitted via pollen in most

10 crops (Maliga, 1993). Plastid transformation in crops has not been widely employed due to the lack of technology. Enhanced expression of selective markers should yield higher transformation efficiencies. The chimeric promoters of the present invention facilitate

15 extension of plastid transformation to agronomically and industrially important crops. Indeed, high-level expression from the PrrnLatpB+DBwt cassette described here resulted in ~25-fold increase in the frequency of kanamycin-resistant transplastomic tobacco lines. More

20 importantly, high levels of marker gene expression following plastid transformation have been obtained in rice, the first cereal species in which plastid transformation has been successful. The results are set forth in Example 8.

25 The following examples are provided to illustrate various embodiments of the present invention. They are not intended to limit the invention in any way.

 The protocols set forth below are provided to facilitate the practice of the present invention.

30

PREPARATION OF CHIMERIC 5' CASSETTES FOR ELEVATED
EXPRESSION OF HETEROLOGOUS PROTEINS IN PLASTIDS OF
HIGHER PLANTS

5 Identification of a potential downstream box in
plastid mRNAs

The presence or absence of downstream box elements in mRNA molecules was determined for the following genes: *psbB* (Tanaka et al., 1987) and *psbA* (Sugita and Sugiura, 1984), photosystem II genes; *rbcL*, encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Shinozaki and Sugiura, 1982); *atpB*, encoding the ATPase β subunit (Orozco et al., 1990); and *clpP*, encoding the proteolytic subunit of the Clp ATP-dependent plastid protease (Hajdukiewicz et al., 1997). Interestingly, most or all of the *PclpP-53* promoter is downstream of the transcription initiation site, therefore the *PrrnLclpP* constructs are assumed to contain two promoters: *Prrn-114* and *PclpP-53*.

10 Transcription initiation sites for these genes were described in references cited above; for nucleotide position of the genes in the plastid genome see Shinozaki et al., 1986.

Initially, it was assumed that the plastid ADB is similar in size and position as the *E. coli* ADB in the 16S rRNA. The *E. coli* ADB is localized on a conserved stem structure between nucleotides 1469 to 1483 (15 nt) that corresponds to nucleotides 1416 and 1430 of the plastid 16S rRNA (Dams et al., 1988; Sprengart et al., 1996). Although in both cases, the ADB is contained in the 16S rRNA penultimate stem, the actual ADB sequence is different in plastids and in *E. coli* (Figure 1B). The N-terminal coding regions of plastid genes *atpB*, *clpP*, *rbcL*, *petA*, *psaA*, *psbA*, *psbB*, *psbD* and *psbE* were

searched for potential DB sequences. The homology search was carried out with a 26 nucleotide sequence centered on the tentative DB region (Figure 1B). The search revealed short stretches of imperfect homology with 5 alternative solutions. Since the position of DB in the mRNA is quite flexible (Etchegaray and Inouye, 1999), we show four potential DB-ADB interactions for *atpB* and *rbcL* in Figure 2A. Two plastid mRNAs were selected to test the role of DB in the translation of plastid mRNAs: 10 1) *atpB* mRNA lacks a SD sequence; and 2) *rbcL* mRNA contains a SD sequence at the prokaryotic consensus. In addition, the phage T7 gene 10 (T7g10) leader was included in the study. This leader has a well-characterized *E. coli* DB sequence (Figure 2B; Sprengart et al., 1996). Additional plastid mRNAs with potential 15 DB sequences shown in Figure 2A are *clpP*, *psbB* and *psbA*.

Experimental strategy to test the efficiency of leader sequences for translation

20 To compare the efficiency of translation from the 5'-UTR of the selected genes, the 5'-UTR was cloned downstream of the strong plastid rRNA operon σ^{70} -type promoter (Prrn-114) (Svab and Maliga, 1993; Allison et al., 1996), which initiates transcription from multiple 25 adjacent nucleotides (-114, -113, -111; Sriraman et al., 1998). The promoter fragments were constructed as SacI-NheI or a SacI-NcoI fragments. Construction of the chimeric promoters using conventional molecular 30 biological techniques is set forth in detail in the next section.

Two constructs were prepared for each 5'-UTR selected: one with (+DB) and one without (-DB) a native

downstream box. It will be obvious from the forthcoming discussion, that the -DB constructs have a synthetic DB provided by the NheI restriction site. The promoters were cloned upstream of the coding region of a kanamycin 5 resistance (neo) gene, which is available on an NheI-XbaI or NcoI-XbaI fragment. For the stabilization of the mRNA, the *rbcL* gene 3'-untranslated region was cloned downstream of *neo* as an XbaI-HindIII fragment. The 10 chimeric *neo* genes can therefore be excised from the pUC118 or pBSIIKS+ plasmids as SacI-HindIII fragments. These source plasmids are listed in Table 1.

Table 1. Salient features of chimeric promoters^a

	Source of 5'-UTR (nucleotides from AUG)	SD	DB	Promoter fragment	pUC118 (U) or pBSIIKS ⁺ (B)	pPRV111A, B
15	atpB (-90/+42)	-	wt	SacI/NheI	pHK10 (U)	pHK30 (B)
	atpB (-90/+6)	-	s	SacI/NheI	pHK11 (U)	pHK31 (B)
	atpB (-90/42)	-	m	SacI/NheI	pHK50 (B)	pHK60 (B)
20	clpP (-53/+48)	-	wt	SacI/NheI	pHK12 (U)	pHK32 (B)
	clpP (-53/+6)	-	s	SacI/NheI	pHK13 (U)	pHK33 (B)
25	rbcL (-58/+42)	+	wt	SacI/NheI	pHK14 (B)	pHK34 (A)
	rbcL (-58/+6)	+	s	SacI/NheI	pHK15 (U)	pHK35 (A)
	rbcL (-58/+42)	+	m	SacI/NheI	pHK54 (B)	pHK64 (A)
30	psbB (-54/+45)	+	wt	SacI/NheI ^d	pHK16 (U)	pHK36 (A)
	psbB (-54/+3)	+	s	SacI/NcoI ^d	pHK17 (U)	pHK37 (A)
35	^b T7g10+DB/Ec (-63/+24) + Ec			SacI/NheI	pHK18 (B)	pHK38 (A)
	^b T7g10+DB/pt (-63/+24) + pt			SacI/NheI	pHK19 (B)	pHK39 (A)
	T7g10-DB (-63/+9) +		s	SacI/NheI	pHK20 (B)	pHK40 (A)
40	psbA (-85/+21)	-	wt	SacI/NheI	pHK21 (U)	pHK41 (A)
	psbA (-85/+3)	-	s	SacI/NcoI ^e	pHK22 (U)	pHK42 (A)
	^c psbA(+GC) (-85/+3)	-		sSacI/NcoI ^e	pHK23 (U)	pHK43 (A)

^aSD+, SD at prokaryotic consensus position; SD-, no SD at prokaryotic consensus position;

DB wt, wild-type; m, mutants; s, NheI site as synthetic DB.

^bEc or pt refers to construct with *E. coli* or plastid DB sequence.

^cpsbA(+GC) indicates addition of GC to the wild-type A at the mRNA 5'-end.

^dIn source gene psbB translation initiation codon is within NcoI site; therefore +DB construct pHK16 has this NcoI site upstream of the NheI site; see Figure 9.

^eTranslation initiation codon is included in NcoI site; NheI site is directly downstream in kan coding region; see Figure 8.

The Prrn promoter fragment is available in plasmid pPRV100A (Zoubenko et al., 1994). The promoters were designed to include sequences between -197 nt and -114 nt upstream of the mature 16S rRNA 5' end. Nucleotide -197 is the 5'-end of the Prrn promoter constructs utilized for these and other studies (Svab and Maliga. 1993; -1 is the first nucleotide upstream of the mature 16S rRNA). The G at the -114 position is one of three transcription initiation sites; the other two are the adjacent C (-113) and A (-111) nucleotides (Allison et al., 1996, Sriraman et al., 1998). The nucleotide at which Prrn transcription would initiate is marked by a filled circle in Figure 3A-D. In most constructs, this is a G (-114) as in the native promoter. In two constructs the G was replaced by an A, as in the *psbA* promoter which is the source of the leader sequence (pHK21, pHK22; see below).

20 DESIGN OF THE 5' LEADER FROM *atpB*

For the *atpB* gene, multiple mRNA 5'-ends were mapped in tobacco leaves including at least four primary transcripts indicating transcription from four promoters and a processed 5'-end 90 nucleotides upstream of the translation initiation codon (Orozco et al., 1990). The terminal nucleotide of the processed *atpB* 5'-end is a G. Therefore, the chimeric PrrnLatpB promoters were designed to initiate transcription at a G, anticipating that the leader sequence of the chimeric transcript will be a perfect reproduction of the processed *atpB* mRNA 5'-end. Out of the *atpB* coding region, 42 and 6 nucleotides are included in the +DBwt and -DB constructs, respectively. The 42 nucleotides include

four potential DB sequences shown in Figure 2A. Two point mutations in the leader sequence were designed to eliminate NheI (T to A) and EcoRI (G to A) restriction sites without affecting the predicted mRNA 5' secondary structure. In the -DB constructs, two codons (6 nucleotides) were retained from the native coding region upstream of the NheI restriction site (GCTAGC sequence) in which the stop codon is out-of-frame (Figure 3A). Eleven silent point mutations were introduced in the DB region of the PrrnLatpB+DBm construct to either minimize the number of base pairs, or to change the nature of base pairing (for example G-C to G-U) (Figure 2A; Figure 3A).

15 **DESIGN OF THE 5' LEADER FROM *clpP***

Two major mRNA 5'-ends of the *clpP* gene were mapped in tobacco leaves (Hajdukiewicz et al., 1997). The terminal nucleotide of the proximal primary transcript is a G. Therefore, the chimeric PrrnL*clpP* promoters were designed to initiate transcription at a G, anticipating that the leader sequence of the chimeric transcript will be a perfect reproduction of the leader transcribed from the P*clp-53* NEP promoter. Out of the *clpP* coding region, 48 and 6 nucleotides are retained in the +DBwt and -DB constructs, respectively. The 48 nucleotides include four potential DB sequences as shown in Figure 2A. In the -DB constructs, two codons (6 nucleotides) were retained from the native coding region upstream of the NheI restriction site (GCTAGC sequence) in which the stop codon is out-of-frame.

DESIGN OF THE 5' LEADER FROM *rbcL*

One primary and one processed mRNA 5'-end were mapped in tobacco leaves for the *rbcL* gene (Shinozaki and Sugiura, 1982). The terminal nucleotide of the 5 processed 5' end is a T. The chimeric PrrnLrbcL promoters were designed to initiate transcription at a G, one nucleotide downstream of the terminal T. Forty-two and 6 nucleotides out of the *rbcL* coding region are included in the +DB and -DB constructs, respectively. The 42 nucleotides include four potential DB sequences as shown in Figure 2A. The one point mutation (G to A) in the leader sequence was designed to eliminate an EcoRI restriction site without affecting the predicted mRNA 5' secondary structure. In the -DB constructs, two codons (6 nucleotides) were retained from the native coding region upstream of the NheI restriction site (GCTAGC sequence) in which the stop codon is out-of-frame. Twelve silent point mutations were introduced into the DB region of the PrrnLrbcL+DBm construct to either minimize the number of base pairs, or to change the nature of base pairing (for example G-C to G-U) (Figure 2A, Figure 3B).

DESIGN OF THE 5' LEADER FROM *psbB*

One primary and one processed mRNA 5'-end for the *psbB* gene were tentatively identified in tobacco leaves (Tanaka et al., 1987). The leader sequence was designed to initiate transcription from the G (-114) of the Prrn promoter, and include the intact secondary (stem) structure assumed to be involved in stabilizing the mRNA. Forty-five and 3 nucleotides out of the *psbB* coding region are included in the +DB and -DB constructs, respectively. The 45 nucleotides include

four potential DB sequences shown in Figure 2A. Since the ATG is naturally included in an NcoI site that is used to fuse the *neo* coding region with the *psbB* leader, no amino acid from the *psbB* coding region is added in the -DB construct.

DESIGN OF THE 5' LEADER FROM *psbA*

One mRNA 5'-end was mapped for the *psbA* gene in tobacco leaves (Sugita and Sugiura, 1984). The terminal nucleotide of the primary transcript is an A. Therefore, the chimeric PrrnLpsbA promoters were designed to initiate transcription at an A, anticipating that the leader sequence of the chimeric transcript will be a perfect reproduction of the leader transcribed from the *psbA* promoter. Twenty-one and 3 nucleotides out of the *psbA* coding region are included in the +DB and -DB constructs, respectively. The 21 nucleotides include the potential DB sequence as shown in Figure 2A. Since the *neo* coding region was linked to the chimeric promoter via an NcoI site which includes the translation initiation codon (ATG), no amino acid from the *psbA* coding region is added in the -DB constructs. This is true of a second -DB promoter, in plasmid PHK23, in which transcription is designed to initiate from the Prrn G (-114) and C (-113) (Figure 3C).

DESIGN OF THE T7 PHAGE GENE 10 LEADER

The T7 phage gene 10 leader (63 nucleotides) was shown to promote efficient translation initiation in *E. coli* (Olins et al., 1988). This leader is used in the *E. coli* pET expression vectors (Studier et al., 1990; Novagen Inc.). The terminal nucleotide at the 5'-end is a G. Therefore, the chimeric PrrnT7g10L promoters were

designed to initiate transcription at a G, anticipating that the leader sequence of the chimeric transcript will be a reproduction of the T7 phage gene 10 mRNA, with the exception of a T to A mutation which was introduced to 5 eliminate an XbaI site. Twenty-four and 9 nucleotides from the T7 phage gene 10 coding region are included in the +DB/Ec (with *E. coli* DB sequence) and -DB constructs, respectively. To compare the efficiency of *E. coli* and plastid DB sequences in plastids, a second 10 +DB promoter was constructed with the tobacco DB sequence (PrrnT7g10L+DB/pt). The native T7g10 leader has an NheI site directly downstream of the translation initiation codon. This NheI site was removed by a T to A point mutation in the +DB constructs (Figure 3D).

15 For introduction into the plastid genome, the chimeric *neo* genes were cloned into plastid transformation vector pPRV111A or pPRV111B. See U.S. Patent 5,877,402, the disclosure of which is 20 incorporated herein by reference. The pPRV111 vectors target insertions into the inverted repeat region of the tobacco plastid genome, and carry a selectable spectinomycin (*aadA*) resistance gene. The sequences of the vectors have been deposited in GenBank (U12812, 25 U12813). The chimeric *neo* gene in vector pPRV111B is in tandem with the *aadA* gene, whereas in vector pPRV111A the chimeric *neo* is oriented divergently. The general outline of the plastid transformation vector with the chimeric *neo* genes is shown in Figures 4A and 4B.

30

CONSTRUCTION OF CHIMERIC Prnn PROMOTERS WITH PLASTID
mRNA LEADERS

The chimeric Prnn promoter/leader fragments were constructed as a SacI-NheI or SacI-NcoI fragments (Table 1, below) by overlap extension PCR (SOE-PCR), essentially as described in Lefebvre et al., (1995). Construction of the Prnn-plastid leader segments is schematically shown in Figure 5. The objective of the PCR-1 step is to 1) amplify the Prnn promoter fragment while 2) adding a SacI site upstream and a seam-less overlap with the specific downstream leader sequence. The reaction contains: 1) a primer (oligonucleotide) to add a SacI site at the 5'-end of the fragment; 2) a suitable template containing the Prnn promoter sequence in plasmid pPRV100A (Zoubenko et al., 1994); and 3) a primer to add on the overlap with the leader sequence at the 3' of the amplified product. The objective of the PCR-2 step is to create the chimeric promoter with DB sequence using: 1) the product of PCR-1 step as a primer; 2) a suitable DNA template containing the specific leader sequence; and 3) primer (oligonucleotide) to include NheI restriction site at the 3'-end of the amplification product. The product of the PCR-2 is the SacI-NheI chimeric Prnn promoter fragment with DB sequence. The objective of the PCR-3 step is to remove the DB sequence while introducing a suitable NheI or NcoI restriction site. The product of PCR-3 is the SacI-NheI or SacI-NcoI chimeric Prnn promoter fragment in which the DB sequence is replaced with the NheI site. The objective of the PCR-4 step is to replace the wild-type DB with a mutant DB. The product of PCR-4 is a SacI-NheI Prnn promoter fragment.

The primers (oligonucleotides) used for the

construction of chimeric promoters are listed in Table 2. The chimeric promoters were obtained by overlap extension PCR using oligonucleotides and DNA templates schematically shown in Figure 5.

5

Table 2.

Oligonucleotides used for the construction of chimeric promoters.

10 #1: 5'-CCCGAGCTCGCTCCCCCGCCGTCTTC-3'
#2: 5'-
CGAATTAAAATAATGTCCGCTTGCACGTCATCGTTAATTCTCCCAGAAATATGCCATCC-3'
15 #3: 5'-CCCGCTAGCCGTGGAAACCCCAGAAC-3'
#4: 5'-CCCGCTAGCTCTCATAATAATAATAATAATATGTC-3'
#5: 5'-TCACTTGAGGTGGAAACGTAACCTCCAGAAATATGCCATCC-3'
20 #6: 5'-CCCGCTAGCTCCTCTCCAGGACTTCG-3'
#7: 5'-CCCGCTAGCAGGCATTAAATGAAAGAAAGAAC-3'
#8: 5'-TAAGAATTTACAACAACAAGGTCTACTCGACTCCCAGAAATATGCCATCC-3'
25 #9: 5'-CCCGCTAGCTTGAATCCAACACTTGCTTTAG-3'
#10: 5'-CCCGCTAGCTGACATAATCCCTCCCTAC-3'
#11: 5'-CAAAGATAAAATAGACACTACGTAACTTATTGCATTGCTCCCAGAAATATGCCATCC-
30 3'
#12: 5'-CCCGCTAGCATCATTCAATAACACGGTATGAACACG-3'
#13: 5'-TTCTAGTGGAAACCGTTGTGGTCTCCCTCCAGAAATATGCCATCC-3'
#14: 5'-CCCGCTAGCCATATGTATATCTCCTCTTAAAG-3'
35 #15: 5'-CCCGCTAGCCTGTCCACCAAGTCATGCTTGCATA-3'
#16: 5'-CCCGCTAGCCAAGGCAGGGCTAGTGATTGCCATATGTATATCTCCTTC-3'
#17: 5'-TTTGTAACTTAAGAAGGAGATACATATGGCAAGCATGACTGGTGG-3'
40 #18: 5'-CTCCTCTTAAAGTTAAACAAAATTATTTCTAGTGGAAACCGTTGT-3'
#19: 5'-CAAAATAGAAAATGGAAGGCTTTGCTCCCAGAAATATGCCATCCC-3'
50 #20: 5'-CAAAATAGAAAATGGAAGGCTTTTCCCAGAAATATGCCATCCC-3
#21: 5'-GGGCCATGGTAAATCTGGTTATTTAATC-3'

#22: 5'-GGGGCTAGCTCTCTCTAAAATTGCAGT-3'
#23: 5'-GAATAGCCTCTCCACCCCA-3'
5 #24: 5'-CCCGCTAGCCGTGGACACCCCACTTCCACTTGTGTCGGGTTATTCTCAT-3'
#25: 5'-CCCGCTAGCTTGAATCCTACTGAGGCTTGTGAGGACTCAT-3'

10

**CONSTRUCTION OF CHIMERIC Prnn PROMOTER/atpB LEADER
SEGMENTS**

PrrnLatpB+DBwt in plasmid pHK10 (Product of PCR-2)

PrrnLatpB-DB in plasmid pHK11 (Product of PCR-3)

15 PrrnLatpB+DBm in plasmid pHK50 (Product of PCR-4)

PCR-1: Oligonucleotides #1, #2 as primers; plasmid pPRV100A (Zoubenko et al., 1994) DNA as template.

PCR-2: Product of PCR-1 step, Oligonucleotide #3 as primers; plasmid pIK79 (see below) DNA as template.

20 PCR-3: Oligonucleotide #1, #4 as primers; Product of PCR-2 step as template.

PCR-4: Oligonucleotide #1, #24 as primers; Product of PCR-2 step as template.

25 Plasmid pIK79 is a Bluescript BS+ phagemid derivative which carries a PvuII/XhoI tobacco plastid DNA fragment between nucleotides 55147-60484 containing the rbcL-atpB intergenic region with divergent promoters for these genes (Shinozaki et al., 1986).

30 **CONSTRUCTION OF CHIMERIC Prnn PROMOTER/clpP LEADER
SEGMENTS**

PrrnLclpP+DBwt in plasmid pHK12 (Product of PCR-2)

PrrnLclpP-DB in plasmid pHK13 (Product of PCR-3)

35 PCR-1: Oligonucleotides #1, #5 as primers; plasmid pPRV100A (Zoubenko et al., 1994) DNA as template.

PCR-2: Product of PCR-1 step, Oligo #6 as primers; tobacco Sal8 ptDNA fragment (Shinozaki et al., 1986) as

template.

PCR-3: Oligonucleotide #1, #7 as primers; Product of PCR-2 step as template.

5

CONSTRUCTION OF CHIMERIC Prnn PROMOTER/rbcL LEADER SEGMENTS

PrrnLrbcL+DBwt in plasmid pHK14 (Product of PCR-2)

PrrnLrbcL-DB in plasmid pHK15 (Product of PCR-3)

10 PrrnLrbcL+DBm in plasmid pHK54 (Product of PCR-4)

PCR-1: Oligonucleotides #1, #8 as primers; plasmid pPRV100A (Zoubenko et al., 1994) DNA as template.

PCR-2: Product of PCR-1 step, Oligonucleotide #9 as primers; plasmid pIK79 DNA (see description of pHK10 above) as template.

15 PCR-3: Oligonucleotide #1, #10 as primers; Product of PCR-2 step as template.

PCR-4: Oligonucleotide #1, #25 as primers; Product of PCR-2 step as template.

20

CONSTRUCTION OF CHIMERIC Prnn PROMOTER/psbB LEADER SEGMENTS

PrrnLpsbB+DBwt in plasmid pHK16 (Product of PCR-2)

25 PrrnLpsbB-DB in plasmid pHK17 (Promoter from pHK16, digested with SacI/NcoI)

PCR-1: Oligonucleotides #1, #11 as primers; plasmid pPRV100A (Zoubenko et al., 1994) DNA as template.

PCR-2: Product of PCR-1 step, Oligo #12 as primers; tobacco Sal8 ptDNA fragment (Shinozaki et al., 1986) as template.

30 PCR-3 was not necessary, since the psbB translation initiation codon is naturally included in an NcoI site. Therefore, the -DB derivative could be obtained by

SacI/NcoI digestion of the PCR-2 step.

CONSTRUCTION OF CHIMERIC Prnn PROMOTER/psbA LEADER
SEGMENTS

5 PrrnLpsbA+DBwt in plasmid pHK21 (Product of PCR-2)
PrrnLpsbA -DB in plasmid pHK22 (Product of PCR-3)
PCR-1: Oligonucleotides #1, #20 as primers; plasmid
pPRV100A (Zoubenko et al., 1994) DNA as template.
PCR-2: Product of PCR-1 step, Oligo #22 as primers;
10 tobacco Sal3 ptDNA fragment (Shinozaki et al., 1986) as
template.
PCR-3: Oligonucleotide #1, #21 as primers; Product of
PCR-2 step as template.

15 PrrnLpsbA(GC) -DB in plasmid pHK23 (Product of PCR-2)
PCR-1: Oligonucleotides #1, #19 as primers; plasmid
pPRV100A (Zoubenko et al., 1994) DNA as template.
PCR-2: Product of PCR-1 step, Oligo #21 as primers;
tobacco Sal3 ptDNA fragment (Shinozaki et al., 1986) as
20 template.

In all of the above, PCR amplification was carried out with AmpliTaq DNA polymerase (Perkin Elmer) or Pfu DNA polymerase (Stratagene) and "stepdown" PCR that utilizes gradually decreasing annealing temperatures was performed (Hecker and Roux, 1996). The exact amplification conditions for the chimeric Prrn::LatpB promoters are given below. The amplification conditions for the remaining chimeric Prrn - plastid leader promoters were calculated according to Hecker and Roux (1996), and differ only in the annealing temperatures. Description of PCR conditions for the construction of the chimeric Prrn promoters with plastid mRNA leaders is

given below; for interpretation of individual steps see scheme in Figure 5.

5

PCR-1 Program: 50 picomoles of both primers per 100 μ l

	1.1 Denature	5 min. at 94 °C	
	2.1 Denature	1 min. at 94 °C	
	2.2 Annealing	0.5 min. at 72 °C	3 cycles
10	2.3 Extension	0.5 min. at 72 °C	
	3.1 Denature	1 min. at 94 °C	
	3.2 Annealing	0.5 min. at 69 °C	3 cycles
	3.3 Extension	0.5 min. at 72 °C	
	4.1 Denature	1 min. at 94 °C	
15	4.2 Annealing	0.5 min. at 66 °C	3 cycles
	4.3 Extension	0.5 min. at 72 °C	
	5.1 Denature	1 min. at 94 °C	
	5.2 Annealing	0.5 min. at 63 °C	3 cycles
	5.3 Extension	0.5 min. at 72 °C	
20	6.1 Denature	1 min. at 94 °C	
	6.2 Annealing	0.5 min. at 60 °C	3 cycles
	6.3 Extension	0.5 min. at 72 °C	
	7.1 Denature	1 min. at 94 °C	
	7.2 Annealing	0.5 min. at 57 °C	20 cycles
25	7.3 Extension	0.5 min. at 72 °C	
	8.1 Extension	10 min. at 72 °C	
	8.2	1 min. at 30 °C	

30 The PCR-2 program was essentially identical to the PCR1 program set forth above with the following modifications: 1) Primers in 100 μ l were the products of 1st PCR reaction, 50 picomoles of the oligonucleotide primer were used; and 2) the annealing temperature in

stepdown PCR was from 67 °C to 52 °C. Accordingly, the following annealing temperatures were used: Step 2.2, 67 °C; Step 3.2, 64 °C; Step 4.2, 61 °C; Step 5.2, 58 °C; Step 6.2, 55 °C; Step 7.2, 52 °C.

5

The PCR-3 and PCR-4 programs were essentially identical to the PCR1 program with the following modification:

1) The annealing temperature in stepdown PCR was from 69 °C to 44 °C. Accordingly, the following annealing temperatures were used: Step 2.2, 69 °C; Step 3.2, 64 °C; Step 4.2, 59 °C; Step 5.2, 54 °C; Step 6.2, 49 °C; Step 7.2, 44 °C. In cases where the yield of the final PCR reaction was too low for efficient cloning, final product was amplified using primers which were used to generate the ends. The final PCR products were digested with the appropriate restriction enzymes (SacI and NheI or SacI and NcoI) and cloned in plasmids pHK2 or pHK3 (see below).

10

15

20

CONSTRUCTION OF CHIMERIC PROMOTERS WITH T7 PHAGE GENE 10 mRNA LEADER SEGMENT

The chimeric Prrn promoter/T7gene10 leader (PrrnLT7g10) fragments were constructed as SacI-NheI fragments (Table 1, below).

25

PrrnLT7g10+DB/Ec promoter in plasmid pHK18

In the absence of a proper DNA template, the PrrnLT7g10+DB/Ec was constructed by employing a modified polymerase chain reaction (Uchida, 1992) in two PCR steps, as schematically shown in Figure 6. The PCR-1A and PCR1B steps generate two fragments in two separate reactions (A and B). The objective of the PCR-1A step is to amplify Prrn promoter fragment while: 1) adding a

SacI site upstream (Oligonucleotide #1 in Table 2); and 2) a seam-less overlap with the specific downstream leader sequence (Oligonucleotide #13 in Table 2) using plasmid pPRV100A (Zoubenko et al., 1994) as DNA template. The objective of the PCR-1B step is to amplify part of the T7g10 leader sequence using overlapping oligonucleotides #15 and #17 in Table 2. The NheI site is introduced in oligonucleotide #15. Both PCR-1A and PCR-1B reactions were carried out by stepdown PCR as described above for the construction of the 10 chimeric Prrn promoters.

PCR-2 reaction generating this chimeric promoter contained:

- 15 a) The products of the PCR-1A and PCR-1B reactions as DNA templates;
- b) Oligonucleotide #18 (0.5 picomole; Table 2) to generate overlapping fragments with products of the PCR-1A and PCR-1B reactions;
- 20 c) Oligonucleotides #1 and #15 (Table 2) for amplification of the final product, 50 picomoles each in 100 μ l final volume.

Promoter was amplified by stepdown PCR, as described for the chimeric Prrn promoters above; the annealing temperatures were between 72 °C to 57 °C.

PrrnLT7g10+DB/pt promoter in plasmid pHK19

The promoter fragment was obtained in one PCR step as shown in Figure 7. The reaction contained:

- 30 a) The product of the PCR-2 reaction generating promoter PrrnLT7g10+DB/Ec in plasmid pHK18 as DNA template; and b) Oligonucleotides #1 and #16 (Table 2), 50 picomoles each in 100 μ l final volume.

Promoter was amplified by stepdown PCR, as described for the construction of chimeric Prrn promoters above; the annealing temperatures were between 72 °C to 52 °C.

5

PrrnLT7g10-DB promoter in plasmid pHK20

The promoter fragment was obtained in one PCR step, which is similar to the PCR-3 step in Figure 5. The reaction contained:

10 a) The product of the PCR-2 reaction generating promoter PrrnLT7g10+DB/Ec in plasmid pHK18 as DNA template; and
b) Oligonucleotides #1 and #14 (Table 2), 50 picomoles each in 100 μ l final volume.

15 Promoter was amplified by stepdown PCR, as described for the chimeric Prrn promoters above; the annealing temperatures were between 72 °C to 52 °C.

20 The final PCR products were digested with the SacI and NheI restriction enzymes and cloned in plasmid pHK3 to obtain plasmids pHK18, pHK19, pHK20.

Construction of chimeric neo genes

Construction of the chimeric promoters was described in the preceding sections. For determining effects on levels of protein accumulation, the promoters were cloned upstream of a kanamycin-resistance encoding construct, consisting of the neo coding region and the 3'-UTR of the plastid *rbcL* gene. Such constructs are available in plasmids pHK2 and pHK3, which carry the same Prrn(L) *rbcL*(S)::neo::TrbcL gene as a SacI-HindIII fragment. Plasmid pHK2 is a pUC118 vector derivative; pHK3 is a pBSIIKS+ derivative. Plasmid maps with relevant restriction sites are shown in Figure 8. DNA

sequence of the neo gene in plasmids pHK2 and pHK3 is shown in Figure 9. Note, that in plasmid pHK2 the neo gene has an EcoRI site upstream of the SacI site (Figure 8). Prrn and TrbCL have been described by Staub and 5 Maliga, 1994; the neo gene derives from plasmid pSC1 (Chaudhuri and Maliga, 1996). The pUC118 and PBSIIKS+ plasmid derivatives which carry the various promoter constructs are listed in Table 1.

To determine the DNA sequence of the promoter 10 fragments, the plasmids were purified with the QIAGEN Plasmid Purification Kit following the manufacturer's recommendations. DNA sequencing was carried out using a T7 DNA sequencing kit (version 2.0 DNA, Amersham Cat. No. US70770) and primer No. #23 in Table 2, which is 15 complementary to the neo coding sequence. These promoter sequences are shown in Figure 3A-D.

Introduction of chimeric neo genes into the tobacco plastid genome

20 Suitable vectors are available for the introduction of foreign genes into the tobacco plastid genome. Such vectors are pPRV111A and pPRV111B, which carry a selectable spectinomycin-resistance (*aadA*) gene and target insertions into the repeated region of the 25 plastid genome (Zoubenko et al., 1994). The chimeric neo genes were cloned into one of these plastid transformation vectors (Table 1) and introduced into the tobacco plastid genome by the biolistic process. From the transformed cells plants were regenerated by 30 standard protocols (Svab and Maliga, 1993). A uniform population of transformed plastid genome copies was confirmed by Southern analysis.

For Southern analysis, total cellular DNA was

prepared by the CTAB method (Saghai-Marof et al., 1984). Two leaves of each transformed plant were homogenized and incubated at 60°C for 30 minutes in a buffer containing 2% CTAB (tetradecyl-trimethyl-ammonium bromide), 1.4 M NaCl, 20 mM EDTA (pH 8.0), 1 mM Tris/HCl (pH 8.0) and 100 mM β -mercaptoethanol. After chloroform extraction, the DNA was precipitated with isopropyl alcohol and dissolved in water or in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). DNA digested with an appropriate restriction enzyme was electrophoresed on 0.8% agarose gel and transferred to nylon membrane using PosiBlot Transfer apparatus (Stratagene). The blots were probed using Rapid Hybridization Buffer and plastid targeting sequences as a probe labeled with random priming (32 P, Boehringer Mannheim Cat No. 1004760).

Plastid transformation was achieved with each of the plasmids listed in Table 1. Exceptions were plasmids pHK41 and pHK42. It appears that NPTII expression with the psbA leader derivatives was so high that the plants were not viable. It follows that these same leaders may be used to advantage when fused with weaker promoters.

Transplastomic lines are designated by Nt (N. tabacum, the species), the plasmid name (for example pHK30) and an individual line number and a letter identifying regenerated plants. For example, the Nt-pHK30-1D and Nt-pHK30-1C plants were both obtained by transformation with plasmid pHK30, are derived from the same transformation event and were regenerated from the same culture. Nt-pHK30-2 plants are derived from an independent transformation event. Normally, several transformed lines per construct were obtained. However, data are shown here only for one: Nt-pHK30-1D, Nt-pHK31-1C, Nt-pHK60-5A, Nt-pHK32-2F, Nt-pHK33-2A, Nt-pHK34-9C,

Nt-pHK35-4A, Nt-pHK64-3A, Nt-pHK36-1C, Nt-pHK37-2D, Nt-pHK38-2E, Nt-pHK39-3B, Nt-pHK40-12B and Nt-pHK43-1C.

5 **Testing mRNA accumulation by RNA gel blot (Northern) analysis**

RNA gel blot analysis was performed to determine steady-state levels of chimeric mRNA in the transplastomic lines. Total leaf RNA was prepared from 10 the leaves and roots of plants grown in sterile culture according to Stiekema et al (1988). RNA (4 μ g per lane) was electrophoresed on 1% agarose gel and transferred to nylon membranes using the PosiBlot Transfer apparatus (Stratagene). The blots were probed using Rapid 15 Hybridization Buffer (Amersham) with a 32 P-labeled neo probe (Pharmacia, Ready-To-Go Random Priming Kit). The neo probe was obtained by isolating the NheI/XbaI fragment from plasmid pHK2. The template for probing the tobacco cytoplasmic 25S rRNA was a fragment which 20 was PCR amplified from total tobacco cellular DNA with primers 5'-TCACCTGCCGAATCAACTAGC-3' and 5'- GACTTCCCTTGCCTACATTG-3'. RNA hybridization signals were quantified using a Molecular Dynamics PhosphorImager, and normalized to the 25S rRNA signal.

25 **Testing NPTII accumulation by protein gel blot (Western) analysis**

Total soluble protein was extracted from the leaves, roots or seeds of transgenic tobacco plants 30 grown in sterile culture. In case of leaves grown in sterile culture, about 200 mg leaf tissue was homogenized in 1 ml of buffer containing 50 mM Hepes/KOH (pH 7.5), 1 mM EDTA, 10 mM potassium acetate, 5 mM

magnesium acetate, 1 mM dithiothreitol and 2 mM PMSF. The homogenate was centrifuged twice at 4 °C to remove insoluble material. Protein concentration was determined using the Biorad Protein Assay reagent kit. Transgenic 5 tobacco plants expressing neo in the plastid genome (Nt-pTNH32-70, Carrer et al., 1993) and wild type plants were used as positive and negative controls, respectively. Proteins were separated in SDS polyacrylamide gels (SDS-PAGE; 15% acrylamide, 6 M urea) 10 and transferred to nitrocellulose membranes using a semi-dry transfer apparatus (Bio-Rad). After blocking non-specific binding sites, the membrane was incubated with 4,000-fold diluted polyclonal rabbit antiserum raised against NPTII (5Prime-3Prime Inc.). HRP- 15 conjugated secondary antibody, diluted 20,000 fold, and ECL chemiluminescence were used for immunoblot detection on X-ray film. NPTII was quantified on the immunoblots by comparison of the experimental samples with a dilution series of commercial NPTII (5Prime-3Prime). 20

EXAMPLE 1

DB sequences enhance protein accumulation from *rbcL* leader; protein accumulation from the *atpB* translation control signals is high but DB-independent

25 The role of DB sequences in mRNA translation was tested using *neo* as the reporter gene. The *neo* gene encodes the bacterial enzyme neomycin phosphotransferase (NPTII) (Beck et al., 1982). The tested *neo* genes have the same promoter (P_{rrn}) and transcription terminator 30 (Tr_{rbcL}), and differ only with respect to the translation control region (TCR) comprising the 5' untranslated region of the mRNA and the coding region N-terminus. Two constructs were prepared with the *atpB* and *rbcL* TCRs.

One construct contained the wild-type TCR, including the processed 5' untranslated region and 42 nucleotides of the coding region N-terminus (PrrnLatpB+DBwt, plasmid pHK30, Figure 4B; PrrnLrbcl+DBwt, plasmid pHK34, Figure 5 4A). The second construct contained silent mutations in the 42-nucleotide segment of the *atpB* and *rbcL* N-terminal coding regions to either eliminate or alter mRNA and rRNA base pairing (PrrnLatpB+DBm plasmids pHK60, Figure 2A and Figure 4B; PrrnLrbcl+DBm, pHK64, 10 Figure 2A and Figure 4A). The silent mutations altered the mRNA sequence without effecting the amino acid sequence. For example, 13 potential base pairs may form between the wild-type *atpB* mRNA and the ADB sequence shown at the bottom in Figure 2A. The 11 silent 15 mutations affect eight base-paring events for this particular ADB-DB interaction. After mutagenesis, there is a possibility for ten base pairing events, most of which are new. The chimeric *neo* genes were introduced into the tobacco plastid genome by homologous targeting 20 using the biolistic approach (Svab and Maliga, 1993; Zoubenko et al., 1994). NPTII and *neo* mRNA levels were then assessed in the leaves of transplastomic plants. Since NPTII in wild-type DB-containing and mutant DB-containing plants has the exact same protein sequence, 25 protein levels in the plants directly reflect the efficiency of mRNA translation. In case of the *atpB* TCR, mutagenesis of DB reduced protein accumulation to ~4% instead of ~7% (Figure 10 and Table 3). In contrast, mutagenesis of *rbcL* DB had a dramatic effect, reducing 30 NPTII accumulation 35-fold. Thus, DB-ADB interaction is very important for translation of the plastid *rbcL* mRNA, but is less important for translation of the *atpB* mRNA.

We also prepared a third construct set with the *atpB* and *rbcL* leaders, but without the native DB (PrrnLatpB-DB, plasmid pHK31, Figure 4B; PrrnLrbcL-DB, plasmid pHK35, Figure 4A). The neo coding region in these constructs is directly linked to the Prrn promoter via a synthetic *Nhe*I restriction site. The *Nhe*I restriction site (GCTAGC) is fully complementary to the ADB region (Figure 2B), therefore it was hoped that it would function as a DB sequence. Utility of *Nhe*I site as an alternative DB could be best judged by NPTII accumulation from the *rbcL* leader, which is highly dependent on DB. High levels of NPTII from the *Nhe*I construct (4.7%) relative to the mutant DB (0.3%) indicate, that linking the coding region via an *Nhe*I site provides a suitable DB for expressing foreign polypeptides (Figure 10, Table 3).

TABLE 3
Levels of NPTII and neo mRNA in tobacco leaves

		SD	DB	NPTII(%)	neo mRNA	NPTII/neo mRNA
20	Nt-pTNH32-70	+	-	2.10±0.33	41.5	5.06
25	Nt-pHK30-1D	(+)	wt	7.02±0.82	70.05±12.33	8.85
	Nt-pHK31-1C	(+)	s	2.52±0.79	100	2.52
	Nt-pHK60-5A	(+)	m	4.03±1.45	91.57±12.76	4.40
30	Nt-pHK32-2F	-	wt	1.17±0.05	49.33±7.76	2.37
	Nt-pHK33-2A	-	s	0.21±0.05	49.55±6.67	0.42
	Nt-pHK34-9C	+	wt	10.83±3.84	48.91±22.65	22.14
	Nt-pHK35-4A	+	s	4.68±1.84	21.41±7.88	21.86
	Nt-pHK64-3A	+	m	0.31±0.15	52.47±4.29	0.59
35	Nt-pHK36-1C	+	wt	2.17±70.97	68.8	3.15
	Nt-pHK37-2D	+	s	2.35±0.05	42.3	5.56

Nt-pHK38-2E	+	Ec	16.39±3.42	47.59±19.06	34.44
Nt-pHK39-3B	+	pt	0.16±0.13	13.12±1.27	1.22
Nt-pHK40-12B	+	s	23.00±5.40	90.27±31.83	25.48
5	Nt-pHK43-1C	(+)	s	0.65±0.28	13.2
					4.92

DISCUSSION

In bacteria, mutagenesis or deletion of the DB reduces translation 2- to 34-fold, depending on the individual mRNA (Etchegaray and Inouye, 1999; Faxén et al., 1991; Ito et al., 1993; Mitta et al., 1997; Sprengart et al., 1996). Furthermore, reliance on the DB increases when the SD sequence is removed (Sprengart et al., 1996; Wu and Janssen, 1996). In our experiments, no variation was made in the *atpB* or *rbcL* 5'UTR, only sequences downstream of the AUG were altered.

Mutagenesis of the *atpB* DB region reduced protein levels ~2-fold. Although the *atpB* mRNA does not have a SD directly upstream of AUG, we speculate that it probably has an alternate mechanism for translation initiation that reduces its dependence on the DB. Alternatively translation initiation may be facilitated by activator proteins as described for *Chlamydomonas* chloroplasts (Rochaix, 1996; Stern et al., 1997). The consequence of DB mutagenesis on *rbcL* translation was a dramatic 35-fold drop in NPTII levels. Accordingly, efficient *rbcL* translation is highly dependent on DB-ADB interactions. Genes in both prokaryotes and eukaryotes show biases in the usage of the 61 amino acid codons and have a tRNA population closely matched to the overall codon bias of the resident mRNA population. Incorporation of synonymous minor codons in the coding region may dramatically reduce translation (Makrides, 1996) and destabilize the mRNA (Deana et al., 1998). A well-

characterized example for minor codons causing reduced expression in *E. coli* are the AGA/AGG arginine codons recognized by the same tRNA which are present at the frequency of 2.6 and 1.6 per thousand codons.

5 Therefore, we have compared codon usage bias and frequency of triplets per 1000 nucleotides in the wild-type and mutagenized *atpB* and *rbcL* DB regions. Since we studied NPTII accumulation in leaves, the values shown in Figure 12 were calculated for the highly expressed 10 *rbcL*, *psaA*, *psaB*, *psaC*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE* and *psbF* photosynthetic genes using the Genetics Computer Group (GCG; Madison Wisconsin) codon frequency program. Codon usage bias and triplet frequency is comparable in the wild-type and mutant DB regions of both *atpB* and 15 *rbcL*. In addition, the mRNAs for the wild-type and mutant DB constructs accumulate at similar levels. Therefore, the dramatic change in NPTII accumulation from the *PrrnLrbcL+DBm* promoter in the Nt-pHK64 line can not be attributed to incorporation of a rare codon in 20 the mutant DB region.

We have shown here that sequences downstream of the translation initiation codon may dramatically affect mRNA translation. Therefore, silent mutations in the DB region of heterologous proteins may significantly 25 improve expression in chloroplasts by increasing complementarity of the mRNA with the plastid rRNA penultimate stem structure.

There are significant differences in NPTII accumulation from neo transgenes with different leaders 30 and the same synthetic DB (Table 3). This indicates that the 5'UTR is an important determinant of translation efficiency. Many data are available supporting the importance of 5'UTR as a target for translational

control in higher plants (Hirose and Sugiura, 1996; Staub and Maliga, 1993; Staub and Maliga, 1994b) and the unicellular alga *Chlamydomonas* (Mayfield et al., 1994; Nickelsen et al., 1999; Sakamoto et al., 1993; Zerges et al., 1997). The data presented herein demonstrate that translation efficiency in plastids is determined by sequences both upstream and downstream of the AUG.

10

EXAMPLE 2

Study of phage T7g10 translation control sequences indicates that the efficient DB in plastids has loose complementarity to ADB

15 Since the actual ADB sequence is different in plastids and *E. coli*, we anticipated (Sprengart et al., 1996; Etchegaray & Inoyue, 1999) that replacement of the *E. coli* DB with a perfect plastid DB (100% DB-ADB complementarity) would enhance translation in plastids.

20 We choose the phage T7g10 translational control region for the study since it has a well-characterized *E. coli* DB. Three Prrn promoter derivatives were constructed. Cassette PrrnLT7g10+DB/Ec consists of Prrn fused with the native T7g10 TCR containing the *E. coli* DB (plasmid pHK38; Figure 2B, Figure 4A). Cassette PrrnLT7g10+DB/pt

25 consists of the Prrn promoter, T7g10 leader and the perfect tobacco DB (pHK39; Figure 2B, Figure 4A). Cassette PrrnLT7g10-DB has the Prrn promoter and T7g10 leader, but lacks the T7g10 DB sequence (pHK40; Figure

30 2B, Figure 4A). The neo coding region in these constructs is directly linked to the Prrn promoter via a synthetic NheI restriction site. The neo genes in the three expression cassettes were introduced into tobacco

plastids by transformation (Svab and Maliga, 1993; Zoubenko et al., 1994) and the leaves of transplastomic tobacco were tested for NPTII accumulation and mRNA levels (Figures 10, 11; Table 3).

5 Surprisingly, NPTII levels from the heterologous T7g10 TCR were higher (Nt-pHK38; ~16%) than the levels obtained from the *rbcL* TCR (Nt-pHK34; ~11%). We expected that incorporation of the plastid DB with 100% complementarity would further enhance NPTII levels.

10 Instead, we found that plants transformed with the construct having the perfect plastid DB (Nt-pHK39) contained NPTII levels 100-fold lower than the plants expressing NPTII from the *E. coli* TCR (Nt-pHK38; Figures 10; Table 3). This result suggests that, unlike in *E.*

15 *coli*, 100% complementarity reduces, rather than enhances translation efficiency. Indeed, none of the highly expressed plastid genes have a perfect DB sequence (Figure 2A). RNA gel blots shown in Figure 11 indicate that Nt-pHK39 plants with the perfect DB contain ~3-fold

20 less *neo* mRNA. Therefore, a contributing factor to lower NPTII levels in these plants appears to be a faster mRNA turnover rate. Furthermore, NPTII expressed from the PrrnLT7g10 derivatives differ by the DB-encoded amino acids at the N-terminus. Therefore, differential protein

25 turnover rates may be part of the reason for differences in NPTII accumulation. The highest yield of NPTII (23%) was obtained with the synthetic, NheI-containing DB cassette.

30

DISCUSSION

This example utilizing the *rbcL* translation control regions reveals that sequences downstream of the translation initiation codon may dramatically affect

mRNA translation. Therefore, silent mutations in the DB region of heterologous proteins may significantly improve expression in chloroplasts by increasing complementarity of the mRNA with the plastid rRNA 5 penultimate stem structure. However, it appears that perfect complementarity is undesirable, as it may accelerate mRNA turnover and reduce the rate of translation. This finding highlights differences in the 10 translation machinery of plastids and *E. coli*, in which perfect complementarity enhances translation (Etchegaray and Inouye, 1999; Sprengart et al., 1996). It is possible, however, that shifting the region of 15 complementarity relative to AUG or targeting a slightly different region of the penultimate stem may facilitate highly efficient translation of mRNAs with a perfectly matched DB.

The T7g10 constructs have one or two relatively rare AGC serine codons (4.7 per 1000, Figure 12), one of which is encoded in the NheI site. This codon is 20 present in the Nt-pHK38 and Nt-pHK40 plants, which contain the highest levels of NPTII. Further improvement may be expected by replacing the AGC with an AGT serine codon.

25

EXAMPLE 3

The *clpP*, *psbB* and *psbA* TCRs have distinct expression characteristics

NPTII accumulation was studied in transplastomic 30 tobacco carrying the PrrnLclpP promoter derivatives. The PrrnLclpP+DBwt (Nt-pHK32-2F) and PrrnLclpP-DB (Nt-pHK33-2A) plants accumulate 1.2% and 0.2% NPTII in their leaves (Figure 10; Table 3). We have found that over-

expression of *clpP* 5'-UTR causes a mutant phenotype manifested as pale green leaf color and slower growth. This phenotype is normalized in older plants. We assume that the primary cause of this mutant phenotype is the 5 lack of ClpP protein, the *clpP* gene product. This mutant phenotype is absent in plants transformed with other 5'UTRs. Therefore we believe, that the mutant phenotype is attributable to competition for a *clpP*-specific nuclear factor. The *clpP* gene has two introns.

10 Preliminary RNA gel blot analysis reveals reduced levels of mature, monocistronic *clpP* mRNA (~30% of wild-type) and accumulation of intron I-containing *clpP* pre-mRNA in the pale-green leaves. Normalization of phenotype coincides with increase of translatable monocistronic 15 *clpP* mRNA to wild type levels. Over-expression of *clpP* 5'UTR therefore may interfere with splicing of *clpP* pre-mRNA.

NPTII accumulation was also studied in transplastomic tobacco carrying the PrrnLpsbB promoter 20 derivatives. The PrrnL psbB+DBwt (Nt-pHK36-1C) and PrrnL psbB -DB (Nt-pHK37-2D) plants accumulate 2.2% and 2.4% NPTII in their leaves (Figure 10; Table 3). Thus, the synthetic DB sequence in case of the *psbB* TCR 25 efficiently replaces the native DB sequence.

Conversely, it may rely on an alternative mechanism for translation initiation.

The Prrn promoter constructs with the *psbA* leader were obtained as described. However, we have been able to introduce only one of them, PrrnLpsbA-DB(+GC) into 30 tobacco plastids in line Nt-pHK43-1C. The Nt-pHK43-1C plants accumulate NPTII at a relatively low level (0.6%; Figure 10, Table 3). It is conceivable that the lack of success in introducing the +DB construct is due to the

dramatically elevated expression level of NPTII which is toxic to the plants.

5

DISCUSSION

NPTII levels obtained from PrrnLclpP+DBwt (Nt-pHK32-2F) promoter are relatively low, only 1.2% of the total soluble protein. However, this promoter is 10 desirable for driving expression of selectable marker genes, as the recovery of transplastomic clones is relatively efficient when the neo gene is expressed from this promoter, as shown in Example 4. Expression of neo from the PrrnLclpP+DBwt promoter does not cause a mutant 15 phenotype in tissue culture. Thus, it is suitable to drive the expression of marker genes, so long as the marker gene is subsequently removed. It appears that competition for a nuclear-encoded factor required for processing the *clpP* introns gives rise to the reduced 20 expression observed. This intron is absent in the *clpP* genes in the monocots rice (Hiratsuka et al., 1989) and maize (Maier et al., 1995). The PrrnLclpP+DBwt promoter therefore may be used to advantage in the transformation 25 of monocots. Furthermore, the level of the trans-factor required for *clpP* intron processing is likely to be expressed at different levels in dicots. We anticipate therefore, that expression of the *clpP* TCR will have no undesirable consequences in other dicot species. It is also possible that the phenotypic consequences of 30 expressing the *clpP* TCR in plastids is a property of the tobacco line, *N. tabacum* cv. Petit Havana utilized herein and is absent in other tobacco lines. This would

make the *clpP* gene TCR a desirable expression tool in both monocots and dicots.

Both *psbB* leader derivatives accumulate NPTII at comparable levels (2.2% and 2.4%, respectively; Table 5). This 5' regulatory region is a good alternative to the most commonly used *rbcL* leader when protein accumulation is required in the ~2% range.

In the past, the *psbA* promoter and leader construct yielded relatively high levels of expression in leaves (2.5% GUS; Staub and Maliga, 1993). Yet these constructs did not contain *psbA* DB elements. The present invention describes the generation of chimeric promoters that are suitable to obtain high-level protein expression while elucidating the regulatory role played by DB sequences. *Prrn* is the strongest known promoter in plastids and consequently provides for high levels of NPTII translation. These elevated levels of NPTII can be toxic to the plant and therefore it is difficult to obtain transplastomic lines with the highest prospective levels of NPTII. An alternative approach involves operably linking the *psbA* leader to a relatively weak promoter. This approach may generate cassettes which are suitable for obtaining relatively high levels of protein accumulation from relatively low levels of mRNA.

25

EXAMPLE 4

NPTII accumulation in roots and seeds

30

Posttranscriptional regulation is an important mechanism of plastid gene expression (Rochaix, 1996; Stem et al., 1997). Therefore, we expected that NPTII accumulation may be tissue-specific due to regulation of

gene expression at the level of mRNA translation. Thus, NPTII accumulation was tested in roots and seeds.

Testing of NPTII accumulation in roots was carried out with a subset of transplastomic lines (Table 4).

5 Roots for protein extraction were collected from plants grown in liquid MS salt medium (3% sucrose) in sterile cultures incubated on a shaker to facilitate aeration. Protein was extracted from the roots with the leaf protocol and tested for NPTII accumulation (Figure 13
10 A). The highest level of NPTII, 0.75%, is found in the roots of plants expressing NPTII from the *clpP* TCR (PrrnL*clpP*+DBwt construct; pHK32). The second highest value, 0.3%, was found in the roots of plants
15 transformed with plasmid pHK38 expressing NPTII from the T7g10 TCR (PrnnLT7g10+DB/Ec promoter). The level of NPTII was about the same, approximately 0.1 %, in roots expressing the recombinant protein from the *atpB* and
20 *rbcL* TCR in pHK30- and pHK34- transformed plants.

Since plastids in the roots are smaller than in leaves, we expected lower levels of NPTII accumulation in the roots than in the leaves. This was true for all the tested roots, except those of the Nt-pHK32 plants. Interestingly, NPTII from the *clpP* TCR accumulated at almost the same level in the roots (0.75%, Table 4) as
25 in the leaves (approximately 1%, Table 3). This is likely attributable to high levels of the neo mRNA in the roots (Figure 13B). Since the *clpP* leader includes the minimal P*clpP*-53 promoter (Sriraman et al., 1998a; NAR 26: 4874) we speculate, that the relatively high
30 mRNA levels are due to activation of P*clpP*-53 in roots. High levels of expression make the *clpP* leader a desirable TCR for protein expression in roots.